© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1717703

AI-Driven Fraud Detection and Financial Security Framework for Saudi Banking Systems

MALIK ASHFAQ UR RAHMAN

Abstract- ***With the advancement towards immediate payments, mobile account opening, cloud-based systems, open banking and data-driven customer journeys in financial services, fraud in digital banking will become increasingly adaptive. This review paper builds a fraud detection and financial security framework using artificial intelligence (AI) for the banking systems of Saudi Arabia in alignment with Vision 2030. The paper asserts that AI can enhance fraud prevention only by incorporating a coordinated approach to machine learning, data management, human-in-the-loop, regulatory compliance and organizational assurances. By following a structured approach to literature review adopted from some Springer-style and systematic review papers published within the last few years, the paper synthesizes scholarly articles related to the use of AI in fraud detection, anomaly detection, privacy-preserving analytics, model management, and cybersecurity of banking institutions. Based on a comprehensive literature analysis conducted during the period 2020 to 2025, five connected capabilities emerge in terms of developing an effective financial security strategy based on AI: trusted data foundation, fraud intelligence modeling, real-time decision controls, investigation, and escalation, and continuous assurance. Through its literature synthesis process, the paper shows how AI helps detect anomalies through several techniques including graph analytics, behavioral biometrics, NLP, ensemble learning, and XAI. The role of fairness testing, model transparency, privacy control in line with the PDPL, and accountability becomes crucial for achieving success.***

Keywords: artificial intelligence, fraud detection, banking security, Saudi banking, Vision 2030, explainable AI, financial crime, model governance.

I. INTRODUCTION

Banks in Saudi Arabia are undergoing a significant wave of digitalisation where customers are using mobile banking, digital wallets, instant payments, digital credit offerings and open banking services. Such trends facilitate the implementation of the Financial Sector Development Program and the entire strategy of Vision 2030, where the government is seeking to establish a competitive and innovative financial sector. However, as the banking sector becomes more convenient, the attack surface for fraud increases. Account takeovers, synthetic identities, phishing, mule accounts, social engineering, payment diversion and digital loan scams are examples of fraud tactics whose impact goes beyond operational issues – they undermine customer trust, threaten financial stability and damage the reputation of regulated entities. Accordingly, artificial intelligence for fraud detection should be understood not as technology alone, but as governance in the modern banking system (Al-Dosari et al., 2022; Ghandour, 2021; Karthik, 2024).

Classical fraud controls rely strongly on rule-based approach, manual investigation process and fraud typology databases. They serve well in many situations, as fraud in banking includes familiar risk factors like unusual transaction volume, suspicious device or repeated login attempts. Nevertheless, traditional methods have difficulties coping with changing fraudsters' behaviour, transaction distribution among accounts or abuse of modern digital interfaces. Artificial intelligence improves upon existing systems by analysing vast amounts of behavioural, transactional, device-related, network or client data. ML models may identify abnormal activity, build clusters based on suspicious relationships, score each transaction's risk level and provide relevant information to investigators through prioritized alerts. AI has proven effective at detecting threats and conducting monitoring in cloud and digital banking environment provided proper governance practices are followed (Choithani et al., 2022; Curzon et al., 2021; Islam Jim et al., 2024).

This topic acquires additional significance in Saudi Arabia, where financial institutions need to innovate and comply with regulations on cyber security, fraud prevention maturity, data management and resilience. The expectation set forth by SAMA in the area of

fraud prevention and counter-cybersecurity emphasizes requirements on risk identification, monitoring, response and countermeasures maturity for regulated entities, whereas the Saudi Personal Data Protection Law imposes additional responsibilities in terms of data handling, privacy, consent and processing. Finally, Vision 2030 calls for fintech development, financial inclusion and improved digital services. For fraud controls to keep up with the pace of innovation, an AI solution must maintain appropriate levels of explainability, fairness and accountability (SAMA, 2024; SDAIA, 2024; Vision 2030, 2024).

The purpose of the review paper is to develop a framework for fraud detection and financial security using AI. The goals are as follows: first, to review available approaches to AI-based fraud detection in banking; second, to highlight governance, privacy and implementation challenges; third, to discuss how AI-based fraud controls can contribute to Saudi financial security strategy; and fourth, to outline a framework that integrates data governance, modelling, decision-making and investigation. The research question is: how can Saudi banking institutions leverage AI in order to improve fraud detection and financial security?

II. LITERATURE REVIEW

Artificial intelligence-based fraud detection is based on the hypothesis that the fraudulent behaviour creates signals during transactions, using different devices, identities, geographic locations, and networks. Logistic regression, random forest, gradient boosting, and neural network are some supervised learning algorithms that classify the transactions in the presence of labelled fraud data. When the new patterns of fraudulent behaviour are not seen before, the unsupervised approaches like clustering, isolation forest, and auto encoder are useful. Deep learning techniques help in understanding complex nonlinear relationships in large volumes of transactions while graphs analytics can reveal coordinated mule networks or interrelationships between the accounts, phones, devices, and beneficiaries. In recent research, these approaches prove to outperform human-driven reviews in terms of speed and scalability when applied to cyber security of banks or digital finance (Almutairi & Nobanee, 2020; Kaloudi & Li, 2020; Karthik, 2024).

Banking fraud detection is different from generic cyber security activities. The mistake of the algorithm could mean the loss of the bank's money in both cases, but allowing a fraud would be acceptable while blocking a customer or delaying salary payment – not. This means that there should be robust model governance and careful operational design in place. The performance metrics in fraud detection are not just accuracy but also include the recall, precision, detection lead time, false-positive cost, explainability, fairness, and ability to aid the investigation process. It seems that the most effective use of the AI technology takes place when the model is embedded in the whole decision lifecycle which starts from alert triaging up to the case closure. Therefore, AI technology in banking could be regarded as decision-making support tool, not as its replacement (Ghandour, 2021; Shneiderman, 2020; Al-Dosari et al., 2022).

Data privacy-preserving approaches become popular recently, given that fraud models require the analysis of customers' personal information. The attached reference paper on AI and privacy-preserving analytics in cloud-based banking shows that AI can contribute to better detection, encryption, automation of regulatory compliance, and secured collaboration of analysts. At the same time, the paper underlines some challenges related to the risks of algorithmic discrimination, false positives, and resources required to adopt these technologies. This perspective is relevant to the study, given that Saudi banks are obliged to comply with PDPL regulations and maintain high-level data governance. Such privacy-preserving technologies like federated learning, differential privacy, tokenization, and secure multi-party computation appear to be popular now as a way to develop or use AI without exposing raw data (Curzon et al., 2021; Liu et al., 2022; Islam Jim et al., 2024).

Another research area is cyber-enabled fraud detection. Fraudulent attacks on banks' customers often begin outside the banking infrastructure through phishing, smashing, fake investment offer,

account impersonation on social media, malware attacks, or credential harvesting. Natural language processing can help to analyze the patterns in phishing letters. Login-device analysis can identify unusual login from new devices, behavioural analysis can catch attempts to create beneficiaries with unusual names, and graph analysis can detect sudden transfers to multiple mule accounts. Another method to investigate the behaviour of a person is behavioural biometrics that analyzes typing patterns, navigation pattern, phone usage patterns, and overall behaviour. This solution can help since the fraudster might have the correct credentials but act differently (Caldwell et al., 2020; Choithani et al., 2022; Karthik, 2024).

The third group of literature sources concerns AI technology adoption for regulatory technology and compliance purposes. AI could be applied to the automation of compliance monitoring and generation of audit logs. It can also find suspicious patterns in transaction flows and detect the patterns needed for the development of anti-money laundering alerts. AI tools could be used by investigators as an additional way to document their decisions about blocking certain transactions. It must be noted that an automated compliance system is only credible when the underlying model can be explained. In case a bank cannot justify its decisions about certain transactions or customers, it would raise questions of regulatory, ethical, and reputational nature. Explainable AI becomes important in fraud detection due to its connection to the decision-making process and accountability (Radanliev & De Roure, 2021; Chen et al., 2022; Chaudhry & Hydros, 2023).

Implementation barriers can be outlined from the literature sources. Firstly, legacy core banking systems might fail to provide transaction feeds in real-time. Second, fraud data could be dispersed among card, payments, digital, and branch operations. Moreover, model training datasets might suffer from bias and the investigation team might not trust the alerts due to noisy output. Finally, smaller banks and fintech companies will likely face challenges in funding and accessing AI talent. Given this evidence, Saudi banks should consider adopting AI technology gradually, starting from the most dangerous types of fraud and expanding to all channels. This paper will combine ideas from the identified literature sources to propose a suitable framework.

III. REVIEW METHODOLOGY

A systematic narrative review will be applied for the current paper that meets the needs for a review paper covering 2020-2025 timeframe. It implies following the logic of recent systematic review models when establishing a topic, defining search boundaries, evaluating relevant sources and developing a framework based on the obtained information. Unlike other systematic review types, the current research did not attempt to conduct a statistical analysis due to differences in datasets and measurements related to fraud, banking, cyber security and privacy issues as well as Saudi financial system transformation. Thus, thematic synthesis is used in the paper that involves integration of evidence obtained from AI, banking risk, cyber security, data privacy, regulation and fintech literature (Tranfield et al., 2020; Snyder, 2020; Islam Jim et al., 2024).

Search criteria include peer-reviewed articles, review papers, regulations, and policies published in the period of 2020-2025. Keywords used for a search include such topics as AI-based fraud detection, banking fraud analytics, machine learning approaches for financial security, anomaly detection, Saudi banks, fintech fraud analytics, explainable artificial intelligence, federated learning techniques, privacy-preserving algorithms, and model risk governance. Criteria for inclusion were relevance to the fields of banking, fraud, cyber security, data privacy, credits or financial security. Papers focused only on non-financial cyber-attacks, solely technical benchmarking and algorithms unrelated to governance and published before 2020 or after 2025 were excluded unless they provided conceptual background. Recent literature, relevancy for Saudi financial sector and practical aspects have been considered during synthesis.

Data extraction involved answering six review questions. First, it identified typologies of fraud covered by AI. Second, it found out appropriate AI methods. Third, it specified required data and privacy controls. Fourth, it outlined decision-making

© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1717703

governance. Fifth, it recognized repeated barriers to implementation. Finally, it analyzed how reviewed evidence could contribute to a framework for Saudi banks. Evidence was divided into such themes as data governance, transaction monitoring, behavioural analysis, model validation, explainability, human review, privacy protection, regulatory alignment, resiliency and vision 2030 benefits.

Figure 1. Review methodology for AI-driven fraud detection in Saudi banking systems

Methodology clarity, banking relevancy, regency, transferability to the Saudi financial institutions and responsible use were among the criteria used to determine the research quality. Claims that lack sufficient explanation, validation or governance mechanisms of the data analysis were evaluated carefully. It is necessary to note that fraud detection algorithms may perform well in datasets, but fail in the real banking environment where customer behavior may change, fraudsters evolve and there is a need to respond quickly within strict timeframes. That is why this review focuses on findings that can be applied in controllable banking fraud investigation process.

IV. FINDINGS AND THEMATIC ANALYSIS

Firstly, one of the most important benefits of AI usage for fraud detection is that the more layers an algorithm has, the greater value it provides. There cannot be a universal algorithm since different types of fraud create unique patterns. For example, transaction-level anomaly detection is useful for unusual amounts, frequency, and time. Device intelligence can find suspicious sessions, emulator utilization, SIM-swap risk, or using new devices. Graph analytics can be used to identify related accounts, beneficiaries, and devices. Natural language processing will facilitate detection of scam messages and frauds. Continuous behavioural biometric identification will ensure constant verification if fraudsters have the credentials to enter. Consequently, multiple methods should be used for fraud detection, and not only one (Kaloudi & Li, 2020; Choithani et al., 2022; Karthik, 2024).

Secondly, AI should work quickly or, at least, provide risk scoring quickly in high-risk channels, such as instant payments and mobile transfers. Algorithms should provide score prior to the confirmation of payment, during the process of beneficiary creation, during login and in the event of drastic changes in customer behaviour. Nevertheless, it is essential to maintain control over this process as high-risk transactions can be verified through step-up authentication, call-back verification, delayed release or even escalated to another employee. Customers can be protected in this manner, and unnecessary friction can be avoided. Table 1 shows AI techniques used for fraud detection in banking.

Table 1. AI techniques for fraud detection in Saudi banking systems

AI technique	Fraud application	Governance requirement	Expected security value
Anomaly detection	Unusual transfers, login behavior and payment velocity	Threshold review and false-positive monitoring	Early warning for unknown fraud patterns
Graph analytics	Mule networks, shared devices and beneficiary links	Data lineage and relationship validation	Detection of coordinated fraud rings
Behavioral biometrics	Session behavior, typing rhythm and device handling	Consent, privacy controls and customer communication	Continuous authentication with lower friction
Natural language	Phishing reports,	Arabic language	Faster identification

© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1717703

processing	complaint text and scam narratives	validation and human review	of social-engineering typologies
Ensemble learning	Combined risk scoring across fraud scenarios	Model validation and performance monitoring	Improved precision and resilience

Thirdly, explainability and human review are pivotal for building trust. Fraud analysts need an explanation about the reasoning behind any flagged transaction. Potential explanations include an unusual beneficiary, the first transaction on a new device, an unusual velocity pattern, an unfamiliar location, transactions beyond a person's average spend and transactions related to known mule accounts. Explanations are also useful when banks explain why further verification is needed to customers. Model explanations should be brief, operationalized and aligned with case management processes. Technical explainability is not sufficient. Human-friendly explanations require understanding by the investigators, compliance officers and management (Shneiderman, 2020; Curzon et al., 2021; Radanliev & De Roure, 2021).

Fourthly, privacy and data governance influence model quality. Models used for fraud detection need clean, reliable and well-integrated data from account management systems, payment rails, cards, call centres, online banking, complaints, device signals and external threat intelligence. Any gaps or poor integration of these data sources elevate model risk. In Saudi Arabia, data lineage, access control, retention policies, consent management where necessary and clear separation between model development and production are essential. The principles of privacy-by-design, such as tokenization, federated learning and differential privacy, are relevant when Saudi banks collaborate with fintechs or exchange threat intelligence without sharing customer data (Liu et al., 2022; Chaudhry & Hydros, 2023; SDAIA, 2024).

Fifthly, fraud governance requires the coordination of operational risk, model risk and financial crime risk. Many Saudi banks have separate teams for fraud operations, cyber security, compliance, data science and products. Such silos hinder quick responses and lessons learned. Therefore, a Saudi banking framework must establish ownership throughout the entire fraud lifecycle. Product teams will design secure customer journeys. Data teams will manage feature pipelines. Model teams will test and validate algorithms. Fraud teams will investigate cases. Compliance teams will oversee regulatory compliance. Internal audits will test governance frameworks. Finally, senior committees will monitor residual risks. Figure 2 summarizes the proposed framework.

Figure 2. AI-driven fraud detection and financial security framework for Saudi banking systems.

Sixthly, Vision 2030 fosters opportunities and responsibilities. On one hand, digitization of financial services, fintech innovations and open banking present opportunities for increased access, better customer experience and greater convenience. On the other hand, they create additional vulnerabilities for fraudsters. In this context, AI-enabled fraud detection can enable Saudi Arabia's vision for a future that includes financial innovation and customer satisfaction in a trusted banking system. Therefore, the proposed framework connects fraud detection to other dimensions, such as trust, stability, innovation and inclusiveness (Vision 2030, 2024; FSDP, 2024; SAMA, 2024).

V. PROPOSED FRAMEWORK

The proposed AI-driven fraud detection and financial security framework consists of five layers. First, trusted data governance sets up the data layer that underlies fraud analytics. Data governance helps

define standards for data quality, customer privacy, lineage, feature ownership and access security. Such a layer is needed since model outcomes hinge upon data completeness and accuracy. Second, fraud intelligence modeling refers to model portfolios that incorporate several types of machine learning techniques such as supervised learning, unsupervised anomaly detection, graph analytics, behavioural biometrics, and natural language processing. The aim here is not about selecting the perfect model but creating a balanced portfolio tailored to specific fraud scenarios.

Third, decision governance converts AI outcomes into structured actions. Decision governance requires setting risk thresholds, requiring human reviews, performing customer verifications, setting up escalation procedures and override controls. Model decisions have to be explainable, auditable and proportional. For example, a suspicious low-value event could require only additional monitoring. A large value payment transfer to a new beneficiary may require step-up verification or temporary freeze of the transaction. Fourth, investigation and security response ensure a seamless link between AI alerts and various financial security actions. These include case management, customer outreach, account protection, incident response, suspicious activities reporting and recovery actions.

Fifth, assurance and continuous learning refer to regular testing of fraud models. As mentioned above, criminal behavior changes. Therefore, banks must test model drifts, false positives, bias issues, investigator feedback, customer complaints and learnings from incidents. Banks should also perform periodic audits and assessments through internal audit and model risk committees that would evaluate documentation, data, model validation and control effectiveness. Table 2 illustrates the proposed framework and governance practices associated with each layer.

Table 2. Framework layers, responsibilities and assurance indicators

Framework layer	Primary owner	Assurance indicators	Vision 2030 contribution
Trusted data governance	Data office, privacy and IT security	Completeness, lineage, access logs and PDPL alignment	Trustworthy digital financial infrastructure
Fraud intelligence modelling	Data science and fraud analytics teams	Precision, recall, drift, bias and explainability	Secure fintech and banking innovation
Decision governance	Fraud operations, compliance and risk committees	Override rate, review time and customer friction	Stable, accountable financial services
Investigation and response	Fraud teams, cyber security and branch operations	Case closure, recovery rate and escalation quality	Customer protection and resilience
Assurance and audit	Internal audit, model risk and board risk committee	Validation records, incident lessons and audit findings	Long-term confidence in digital banking

This fraud prevention framework is suitable for Saudi banks as it addresses three local priorities. First, it fosters the development of a counter-fraud maturity level compliant with the policies of SAMA. It does so by providing fraud risks to be assessed, measured, governed and continuously improved. Second, the framework helps Saudi banks comply with the PDPL requirements related to privacy. Specifically, it enables data protection by incorporating privacy measures into machine learning and model development. Third, this framework advances digital banking capabilities within the vision of Saudi Arabia.

VI. DISCUSSION

According to the review, AI may provide a powerful means for detecting fraud in banks, yet technology cannot solve banking fraud problems alone. Banking fraud is a dynamic social phenomenon that is adaptable, cross-channel and relies on the exploitation of human trust. As such, AI-based models should be integrated into processes of educating customers about phishing, verifying

customer identities, protecting financial products from misuse and conducting timely investigations. Moreover, Saudi Arabia faces particular challenges since its citizens actively use digital channels and fintech services. Thus, an effective solution should be balanced so as not to harm trust or scare customers away from using digital banking services.

One of the key implications of the review is that banks should evolve from alerts to decision intelligence. Most financial institutions generate thousands of alerts daily. Still, an abundance of alerts does not necessarily equate to improved security because badly designed alerts will overload investigators and result in missed cases. By contrast, an AI-powered framework should filter out cases, rank them in terms of risk, explain the drivers of risk and suggest the next step to be taken. The results obtained after investigating fraud or false alarms should be used for training the machine-learning algorithms, which would foster continuous improvement aligned with mature governance standards (Al-Dosari et al., 2022; Ghandour, 2021; Radanliev & De Roure, 2021).

Another implication is that AI fraud detection should be seen as a model risk activity. The objective of model developers may be to achieve high efficiency metrics. Yet, the board and executive officers must ensure that algorithms are lawful, fair, resilient, auditable and explainable. In Saudi Arabia, this issue must be approached with particular attention paid to explainability as customer communication should be carried out in Arabic. Thus, banks should invest in developing localised AI solutions and performing relevant analytics in Arabic (SDAIA, 2024). The process of managing model risks should involve independent validation, stress-testing, scenario-testing, bias review and post-implementation monitoring.

According to the review, collaboration is another aspect of fighting against fraud that is critically important for Saudi banks. Fraud schemes are often collaborative and spread across multiple banks, wallets and fintech services. Hence, a single institution can see only partial patterns and needs to rely on information from other parties. Collaboration in the form of privacy-preserving analytics, joint typology identification and regulator-sponsored exchange of intelligence will help in improving fraud detection without compromising privacy. Methods like federated learning can benefit both banks and fintech companies, provided that proper data-use agreements have been signed (Chaudhry & Hydros, 2023; Liu et al., 2022).

VII. CONCLUSION AND RECOMMENDATIONS

In summary, the current review paper provides an overview of how AI can help detect fraud and protect Saudi banking systems in light of Vision 2030. The analysis shows that AI algorithms can boost banking security by performing real-time anomaly detection, analysing networks based on graph analytics, acquiring behavioural intelligence, using NLP methods, employing ensemble learning techniques and providing explanations via explainable scoring. Nevertheless, AI-powered fraud detection cannot be effective without proper data, governance practices, and continuous assurance. The resulting framework links technical solutions and governance components that help in making safe decisions.

It is recommended that Saudi banks implement the proposed framework via a phased approach to fraud prevention. The first phase consists in assessing typical fraud schemes and identifying the extent to which banks are prepared to detect them. The second phase includes piloting machine learning algorithms to detect typical fraud cases, such as account takeovers or suspicious mule accounts. In the third phase, the alerts produced by machine learning algorithms are included into the process of fraud investigation and customer identity verification. Finally, during the fourth phase, model risk governance and bias testing are introduced. The fifth phase involves setting up collaboration with fintech partners, regulators and other banks while protecting customer privacy.

For regulators and policy makers, it is suggested that they maintain the focus on counter-fraud maturity, responsible AI, privacy-by-design and secure open banking. For banks, cross-functional ownership of the project that includes representatives of fraud operations, cyber security, compliance, data science,

legal, product and customer experience departments is necessary. For technology providers, it is advised that they design algorithms that allow for transparency, Arabic-language threat intelligence and thorough documentation. Future research is needed in order to experimentally test the framework using Saudi banking cases, analyse models' performance in various types of fraud and investigate customer attitudes towards AI-powered fraud prevention.

VIII. ROADMAP FOR PRACTICAL IMPLEMENTATION

First of all, a diagnostics phase is necessary prior to implementing any machine learning model. A Saudi bank should identify its unique typologies of fraud, digital channels involved, types of customers, payment instruments, source of complaints, as well as detection mechanisms currently in place. By identifying where losses happen, where friction arises for customers, and in which areas existing rules generate too many false positives, banks can establish a risk-based portfolio of use cases. This way, a generic artificial intelligence initiative becomes a practical application of algorithms that solves specific operational challenges in fraud detection.

Secondly, data readiness is a necessity. Since AI-based fraud detection depends on having a well-managed environment of features, the problem is that most banking databases are scattered across various systems designed for transactions, rather than analytics. For that reason, a bank needs to build a feature catalogue to specify what variables are approved, their refresh interval, owner, lineage, privacy level, and permitted usage. Examples of features might include change in device, login velocity, beneficiary age, transaction history, complaint indicators, failed authentication, customer segment, network relation. Each of the features must be tested for stability and bias, especially when demographic, geographic or financial variables may lead to biased decisions.

Thirdly, banks need to develop and validate a portfolio of fraud detection models that include supervised, unsupervised, graph and natural language models. While supervised models are used for detecting known typologies of frauds, unsupervised ones are used to detect emerging anomalies, graph models expose mule relations between accounts, and natural language models help to analyze scammers' messages as well as customers' reports. Validation must involve back-testing, out-of-time testing, scenario analysis and adversarial testing. It is insufficient to approve a model just because it detects past incidents, but also because it is operationally useful, interpretable by investigators, stable when customer behaviour changes, and monitorable post-deployment.

Fourthly, deaccessioning process should incorporate the idea of proportional response to each alert, whether low-, medium- or high-risk. Low-risk alerts might go to a monitoring queue, while medium-risk alerts require additional confirmation. High-risk alerts might warrant a delay in transactions or even an automatic escalation to case management platform and protection of an account. In any case, it is always necessary to have a person authorized to make a decision, and such decision cannot rely exclusively on automated tools, at least in cases where customers lose access to products or experience significant financial impacts. Decision policy has to define who can override a model, what evidence is needed, how to communicate with the customers and resolve complaints.

Fifthly, fraud prevention model needs to be integrated into a bank's workflows. Otherwise, it will not have an effect on operational processes or fraud losses. This means that alerts generated by a model must be processed in case management system along with recommendations on how to handle each case, priority, context and additional information for investigating. Moreover, investigators should be able to document their decisions, flag false positives, write narrative comments, and update models based on their findings. On the other hand, product owners should be aware of fraud trends across customer journeys, and compliance teams – about escalation of suspicious cases.

Step six is assurance, audit and continuous improvement. Frauds models need to be checked for issues such as drift, degradation, bias, the burden on false positives and developing new typologies. Internal audit should test for documentation, controls

around data, validation, access and management reporting. Model risk committees should ensure outputs stay aligned with the risk profile and customer treatment. Banks should conduct post-mortem reviews following any major fraud incidents and revise any rules, features, consumer notifications and analyst training based on these reviews. Criminals learn and adapt quickly, and the bank's cycle of improvement must outpace theirs.

Step seven is ecosystem collaboration. Banking security in Saudi Arabia is better when multiple banks, Fintech firms, payment processors, telecom operators and regulators work together to share intelligence. Many fraud patterns flow between companies, and each one operates in a silo. Collaboration might include typology warnings, secure data exchange procedures, federated machine learning, consumer education programs and joint response planning. All of this needs to be done with strict adherence to privacy principles, including tokenisation, aggregation and appropriate controls for sensitive information. Thus, the framework creates a secure financial ecosystem rather than individual company security alone.

Finally, there is capability building. Vendor solutions should not rely upon fully for fraud governance, since fraud involves a local understanding of customers, communication in Arabic and emerging domestic scams. Any good programme needs to develop the skills of fraud analysts for interpreting data, data scientists who understand banking risks, compliance teams on model governance and customer facing staff on safe reporting practices. Dashboards for senior management should clearly present how AI is reducing losses, improving customer experience, ensuring compliance and creating strategic certainty. Capability building also means procurement standards. For example, vendors should be evaluated on explain ability, integration, documentation, cyber security, data residency, local support and knowledge transfer. With these elements in place, AI becomes a corporate capability, not just technology purchased from another firm. This is important for Vision 2030, since the secure development of digital finance depends on Saudi banks innovating in a secure and customer-focused manner.

Implementation should be measured through a range of balanced indicators, not simply those related to technical deployment. Indicators to measure should be successful cases of fraud avoided, false positive reduction, improved customer verification times, increased efficiency for investigators, increased recovery value, investigator acceptances, drift events managed, audit findings resolved and privacy issues prevented. However, banks need also to monitor the level of customer harm caused, even by the avoidance of fraud, since lack of clear communications and long delays create consumer stress regardless of the outcome. Therefore, the reporting framework needs to capture both the financial gains and customer protection impacts of AI. This is critical for board oversight and allows regulators and management to determine between experimentation and sound governance of financial security.

As open banking, instant payments and embedded finance grow, the road map needs to be flexible. New partners will bring new interfaces, data dependencies and customer journeys. At the moment of any material change, the bank should review their fraud risk and impact to their models. Customer communications should also be tested and improved. In this way, flexibility is brought into the governance framework, and it remains responsive to innovation but continues to retain its disciplines.

REFERENCES

[1] Al-Dosari, K., Fetais, N., & Kucukvar, M. (2022). Artificial intelligence and cyber defense system for banking industry: A qualitative study of AI applications and challenges. Cybernetics and Systems, 55(2), 302-330.

[2] Almutairi, M., & Nobanee, H. (2020). Artificial intelligence in financial industry. SSRN Electronic Journal.

[3] Bouteraa, M., Raja Hisham, R. R. I., & Zainol, Z. (2022). Challenges affecting bank consumers' intention to adopt green banking technology in the UAE: A UTAUT-based mixed-methods approach. Journal of Islamic Marketing, 14(10), 2466-2501.

[4] Caldwell, M., Andrews, J. T. A., Tanay, T., & Griffin, L. D. (2020). AI-enabled future crime. Crime Science, 9(1), 1-13.

[5] Chaudhry, U. B., & Hydros, A. K. M. (2023). Zero-trust-based security model against data breaches in the banking sector: A blockchain consensus algorithm. IET Blockchain, 3(2), 98-115.

[6] Chen, J., Henry, E., & Jiang, X. (2022). Is cybersecurity risk factor disclosure informative? Evidence from disclosures following a data breach. Journal of Business Ethics, 187(1), 199-224.

[7] Choithani, T., Chowdhury, A., Patel, S., Patel, P., Patel, D., & Shah, M. (2022). A comprehensive study of artificial intelligence and cybersecurity on bitcoin, cryptocurrency and banking system. Annals of Data Science, 11(1), 103-135.

[8] Curzon, J., Kosa, T. A., Akalu, R., & El-Khatib, K. (2021). Privacy and artificial intelligence. IEEE Transactions on Artificial Intelligence, 2(2), 96-108.

[9] Financial Sector Development Program. (2024). Annual report and delivery plan updates under Saudi Vision 2030. Riyadh: Vision 2030.

[10] Ghandour, A. (2021). Opportunities and challenges of artificial intelligence in banking: Systematic literature review. TEM Journal, 10(4), 1581-1587.

[11] Iman, N., Nugroho, S. S., Junarsin, E., & Pelawi, R. Y. (2023). Is technology truly improving the customer experience? Analysing the intention to use open banking in Indonesia. International Journal of Bank Marketing, 41(7), 1521-1549.

[12] Islam Jim, M. M., Hasan, M., & Khatun Munira, M. S. (2024). The role of AI in strengthening data privacy for cloud banking. Frontiers in Applied Engineering and Technology, 1(01), 252-269.

[13] Kaloudi, N., & Li, J. (2020). The AI-based cyber threat landscape: A survey. ACM Computing Surveys, 53(1), 1-34.

[14] Karthik, M. (2024). Cybersecurity threats in banking: Unsupervised fraud detection analysis. International Journal of Science and Research Archive, 11(2), 915-925.

[15] Liu, B., Pavlou, P. A., & Cheng, X. (2022). Achieving a balance between privacy protection and data collection: A field experimental examination of a theory-driven IT solution. Information Systems Research, 33(1), 203-223.

[16] Radanliev, P., & De Roure, D. (2021). Alternative mental models for artificial intelligence privacy and security. IEEE Technology and Society Magazine, 40(3), 37-47.

[17] Radanliev, P., De Roure, D., Nurse, J. R. C., Montalvo, R. M., Cannady, S., Santos, O., & Burnap, P. (2021). Cyber risk from IoT technologies in the supply chain. Sensors, 21(6), 2144.

[18] Saudi Central Bank. (2024). Counter-Fraud Framework. Riyadh: SAMA.

[19] Saudi Central Bank. (2024). Adherence to the Personal Data Protection Law and data governance policies, regulations and rules. Riyadh: SAMA.

[20] Saudi Data and AI Authority. (2024). Personal Data Protection Law and implementing regulatory guidance. Riyadh: SDAIA.

[21] Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe and trustworthy. International Journal of Human-Computer Interaction, 36(6), 495-504.

[22] Snyder, H. (2020). Literature review as a research methodology: An overview and guidelines. Journal of Business Research, 104, 333-339.

[23] Tranfield, D., Denyer, D., & Smart, P. (2020). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. British Journal of Management, 31(1), 173-194.

[24] Truby, J., Brown, R., Dahdal, A., & Ibrahim, I. (2020). Blockchain, climate damage, and death: Policy interventions to reduce the

carbon emissions, mortality, and net-zero implications of non-fungible tokens and bitcoin. Energy Research & Social Science, 88, 102499.

[25] Vision 2030. (2024). Financial Sector Development Program and national transformation progress. Riyadh: Government of Saudi Arabia.

[26] Wamba-Taguimdje, S. L., Wamba, S. F., Kamdjoug, J. R. K., & Wanko, C. E. T. (2020). Influence of artificial intelligence on firm performance: The business value of AI-based transformation projects. Business Process Management Journal, 26(7), 1893-1924.

[27] Wang, H., Wang, Y., & Wang, X. (2021). Artificial intelligence in financial fraud detection: A systematic review. Journal of Risk and Financial Management, 14(10), 497.

[28] Xie, L., & Wang, S. (2023). Privacy-preserving machine learning for financial risk control. Expert Systems with Applications, 213, 118957.

[29] Zhang, Y., Li, T., & Liu, Y. (2023). Federated learning for fraud detection in banking: Privacy, performance and governance. IEEE Access, 11, 56631-56645.

[30] Zhou, Z., Chen, X., & Li, Y. (2025). Explainable artificial intelligence for financial crime detection: A review and research agenda. Information Systems Frontiers, 27(1), 155-178.